**Project Ideas:**

1. **Emotion Detection in Text**: Develop a model to classify text based on the emotion conveyed. This could be applied to social media posts, customer reviews, or any text data.
2. **Personality Prediction**: Use text data to predict the personality of the author. This could be based on the Big Five personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism.
3. **Mental Health Monitoring**: Analyze social media posts or other text data to identify signs of mental health issues such as depression or anxiety.
4. **Psycholinguistic Analysis**: Investigate how language and cognition are related. For example, you could look at how complexity of language changes with age, or how language use might be related to cognitive decline.
5. **Chatbot for Mental Health Support**: Develop a chatbot that can provide basic mental health support, like active listening and providing positive affirmations.
6. **Analyzing Therapeutic Conversations**: Therapeutic conversations could be analyzed to understand patterns, measure progress, and even predict outcomes.

Obtained from GPT:

Language Complexity and Age: Analyze text data from different age groups to see how language complexity changes with age. For example, you might hypothesize that older individuals use more complex sentence structures or have a larger vocabulary. You could test this hypothesis by collecting text samples from various age groups and calculating metrics like average sentence length, vocabulary diversity, etc.

Cognitive Decline and Language Use: Research has shown that changes in language use can be an early indicator of cognitive decline, including conditions like Alzheimer’s disease. With the right dataset (and necessary ethical approvals), you could explore this relationship. For example, you might look at how the use of certain words or phrases, or the complexity of sentences, changes over time for individuals with cognitive decline.

Language and Memory: Memory plays a crucial role in our ability to use language. You could explore this relationship by analyzing text data from memory recall experiments. For example, you might look at how the amount of detail or the structure of recall changes based on how much time has passed since the event.

Language Processing Speed: Psycholinguistics also involves the study of how quickly and efficiently we can process language. While this might be harder to study using text data alone, you could look at related measures. For example, you might analyze how quickly individuals respond to different types of questions, or how reading speed varies with text complexity.

**Research Papers:**

[AI-based personality prediction for human well-being from text data: a systematic review | Multimedia Tools and Applications (springer.com)](https://link.springer.com/article/10.1007/s11042-023-17282-w)

[2307.14385v2.pdf (arxiv.org)](https://arxiv.org/pdf/2307.14385v2.pdf)

**Datasets:**

Language and cognition: <https://www.kaggle.com/datasets/manishkc06/human-memory-and-cognition>

<https://www.kaggle.com/datasets/tapakah68/impressive-dataset>

Mental Health:

<https://www.kaggle.com/datasets/vhonghoavin/depression-dataset-after-preprocessing>

<https://www.kaggle.com/code/ahmedashrafahmed/medical-chatbot-using-gpt-2>

**Models:**

BERT(Variants), GPT-4, PaLM, FLAN-T5, LLaMA, Alpaca

Problem Statement

Script:

Script:

We aim at analyzing the patterns in student’s dropout rates. For this we have obtained records of over 4000 students. We will be analyzing their academic record, their backgrounds, socio economic factors, and demographics. We observed that every 1 in 3 students drops out of highschool. These numbers are alarming and we aim to find a solution for this. We will be visualizing the data with the following plots. Chord Diagrams and Pie Charts to visualize categorical data. MDS plot and radar plot to identify correlations between various numerical features. Finally, once we have those correlations, we can test them on the scatter plot and draw some actionable insights. We aim at finding the root causes of student dropouts and thus propose solutions to it.